

THE NAS KERNEL BENCHMARK PROGRAM

David H. Bailey and John T. Barton

Numerical Aerodynamic Simulations Systems Division
NASA Ames Research Center
June 13, 1986

SUMMARY

A benchmark test program that measures supercomputer performance has been developed for the use of the NAS (Numerical Aerodynamic Simulation) Projects Office at NASA Ames Research Center. This benchmark program is described in detail and the specific ground rules for running the program as a performance test are discussed.

INTRODUCTION

A benchmark test program has been developed for use by the NAS program at NASA Ames Research Center to aid in the evaluation of supercomputer performance. This program consists of seven Fortran test kernels that perform calculations that are typical of Ames supercomputing. It is expected that the performance of a supercomputer system on this program will provide an accurate projection of the performance of the system on actual NAS program computer codes. This paper describes the test program in detail and lists the specific ground rules that have been established for running the program as a performance test.

PROGRAM DESCRIPTION

The NAS Kernel Benchmark Program consists of approximately 1000 lines of Fortran code, organized into seven separate tests. Each individual test consists of a loop that iteratively calls a certain subroutine. These subroutines were chosen after review of many of the calculations currently being performed on Ames supercomputers and by recommendations from a number of Ames scientists and programmers, particularly those working on computational fluid dynamics problems. In most cases, these subroutines have been extracted from actual programs currently in use, and they have been incorporated into the NAS Kernel Benchmark Program with only minor changes. Thus it is felt that these test kernels are a representative cross section of expected NAS program supercomputing, and the performance of a computer system (both its hardware and its Fortran compiler) on these tests should be a reliable predictor of the actual system performance on NAS user programs.

The seven selected programs all emphasize the vector performance of a computer system. Almost all of the floating-point operations indicated in these Fortran subroutines are contained in loops that are computable by vector operations, provided that the Fortran compiler of the computer system being tested is sufficiently powerful in its vectorization analysis, and provided that the hardware design of the computer includes the necessary vector instructions. Most serious supercomputer programs currently in use at Ames are fairly highly vectorized, and it is expected that programs to be developed in the future will virtually all be designed to effectively use the vector processing capabilities of supercomputers. Some programs that have substantial scalar processing will continue to be used, but it is expected that their numbers will decline as algorithms and codes that are more suitable for vector processing are developed. Another reason for emphasizing vector performance in these benchmark kernels is that it is not very meaningful to average, even in a harmonic average sense, the performance of a supercomputer on a scalar code with its performance on a vector code.

This program not only tests the hardware execution speed of a computer, but it also tests the effectiveness of the Fortran compiler. It is clear that a phenomenally fast hardware

design is worthless unless it is coupled with a Fortran compiler that can fully utilize the advanced hardware design. Furthermore, it is becoming increasingly clear that vectorization and other optimizations must either be completely automatic or be very easy to direct. If effective utilization of a computer requires massive redesign of otherwise well-written, standard Fortran-77 code, or if a high level of performance is possible only by considerable human intervention, then the actual usable power of the computer is severely reduced.

The seven test kernels of the NAS Kernel Benchmark Program have, for the most part, been developed quite recently. As a result, they represent Fortran programs that have been designed and written for modern vector computation, as opposed to the somewhat dated code that is used for other popular benchmark programs. It might be argued that there is some inherent bias in the test towards the Cray computers, since most of these kernels were written on a Cray X-MP. However, substantial care was exercised in the selection of these kernels to insure that none of them had any constructs that would unduly favor the Cray line. As much as possible, subroutines were selected that were merely straightforward Fortran code, intelligently coded with loops that are capable of being executed with vector operations, but otherwise neutral towards any particular machine. In fact, in the process of selecting these kernels for testing, it was discovered that some of them actually caused unforeseen difficulties for the Cray compiler. Nevertheless, they were left in the test suite to maintain objectivity.

Performance is measured by the NAS Kernel Benchmark Program in MFLOPS (millions of floating-point operations per second). The precise number of floating-point operations for the various functions used in the test kernels is shown in Table 1. These numbers are based on actual counts of 64-bit floating-point operations in published algorithms.

It should be noted that this program only measures MFLOPS rates. Disk I/O, operating system efficiency, and other important factors of overall performance are not measured by this benchmark program. Also, several of the test subroutines perform a significant amount of memory move, integer, and logical operations, none of which is included in the floating-point operation count.

The following is a description of the seven proposed Fortran test kernels. Other features are summarized in Table 2.

1. MXM – This subroutine performs the usual matrix product on two input matrices. The subroutine employs a four-way unrolled, outer product matrix multiply algorithm that is especially effective for most vector computers. See [1] for a discussion of this algorithm.
2. CFFT2D – This test performs a complex radix 2 FFT on a two dimensional input array, returning the result in place. The test kernel actually consists of two subroutines that perform FFTs along the first and second dimension of the array, respectively, taking advantage of the parallel structure of the array. See [2] for a discussion of the FFT algorithm used.

Table 1: Floating-point Operation Counts

FIRST ARGUMENT	FUNCTION	SECOND ARGUMENT	FLOATING PT. OPS.
Real	+	Real	1
Real	-	Real	1
Real	*	Real	1
1	/	Real	2
Real	/	Real	3
Real	**	2	1
Real	**	Real	45
Complex	*	Real	2
Complex	/	Real	4
1	/	Complex	7
Real	/	Complex	9
Complex	+	Complex	2
Complex	-	Complex	2
Complex	*	Complex	6
Complex	/	Complex	13
Real	SQRT		12
Real	EXP		18
Real	LOG		25
Real	SIN		25
Real	ATAN		25
Complex	ABS		15
Complex	EXP		70
Complex	LOG		65

Table 2: Kernel Features

FEATURE	KERNEL						
	1	2	3	4	5	6	7
Two dimensional arrays	X	X			X	X	X
Multidimensional arrays			X	X			X
Dimensions with colons			X				
Integer arrays		X			X	X	
Integer functions in indices					X	X	
IF statements in inner loops						X	
Scientific function calls		X	X		X	X	
Complex arithmetic		X			X	X	
Complex function calls					X	X	
Inner loop memory strides	1	1	1	1	1	1	128
		2	4	2	2		
		256		750	500		
				900			
Inner loop vector lengths	256	128	250	28	5	100	128
		256			100	500	
					500	1000	

3. CHOLSKY – This subroutine performs a Cholesky decomposition in parallel on a set of input matrices, which are actually input to the subroutine as a single three-dimensional array.
4. BTRIX – This kernel performs a block tridiagonal matrix solution along one dimension of a four dimensional array.
5. GMTRY – This subroutine sets up arrays for a vortex method solution and performs Gaussian elimination on the resulting array. This kernel is noted for a number of loops that are challenging to vectorize.
6. EMIT – Also extracted from a vortex code, this subroutine creates new vortices according to certain boundary conditions.
7. VPENTA – This subroutine simultaneously inverts three matrix pentadiagonals in a highly parallel fashion.

In each of the above test subroutines, the input data arrays are filled by a portable pseudorandom number generator in the calling program. This feature insures that all computers running the NAS Kernel Benchmark Program will perform the required calculations

on the same numbers. It also permits the output results to be checked for accuracy. Each of the seven tests is independent from the others – none depends on results calculated in a previous test program. Thus program alterations to improve the execution speed of one of the test kernels may be made without fear of affecting the other kernels.

GROUND RULES FOR PERFORMANCE TESTING

Worlton's recent article [3] pointed out some of the difficulties that are involved in supercomputer performance testing. Most of these problems are a result of the lack of well-defined controls on these tests. For instance, in some recent test results, one vendor was apparently allowed to perform some minor tuning and insertion of compiler directives, whereas the other was not. In other cases confusion has resulted from researchers not carefully noting exactly which version of a vendor's compiler was being used in their tests. Some vendors have claimed amazingly high performance rates for their computers, which, upon closer analysis, have been achieved only by massive recoding of the test kernels and by the usage of assembly code. As a result of these difficulties, many of the recent comparisons of supercomputer performance have degenerated into shouting matches that have generated more heat than light.

In consideration of such problems, some strict ground rules have been established for using the NAS Kernel Benchmark Program to evaluate supercomputer performance. Also, four levels of tests have been defined, so that the effects of varying amounts of tuning may be assessed. These different levels will also enable the NAS program to differentiate the performance of the hardware from that of the compiler. If the compiler is truly effective, then a relatively small amount of tuning should be sufficient to achieve close to the full potential of the hardware. The four test levels are defined as follows:

1. Level 0 (“dusty deck”): For this test, the NAS Kernel Benchmark Program must be run without any changes to improve performance. If any alterations are required for compatibility purposes (for example, to define the timing function), they must be made by NAS program personnel.
2. Level 20 (“minor tuning”): For this test, a few minor alterations may be made to the code to enhance performance. These changes may include, for example, compiler directives to assist the compiler's vectorization analysis or changes to array dimensions to avoid disadvantageous memory strides. No more than 20 lines of code in the entire program file may be inserted or modified.
3. Level 50 (“major tuning”): For this test, more extensive modifications may be made to the code to enhance performance. For example, some loops may be rewritten to avoid constructs that cause difficulties for the compiler or the hardware. A total of up to 50 lines of the program file may be inserted or modified for this test.

4. Level 1000 (“customized code”): For this test, large scale coding changes are allowed to improve performance. Entire subroutines may be rewritten to avoid difficult constructs. There is no limit to the number of lines of code that may be inserted or modified.

For all four levels of tests, any modifications made to the program code must conform to the ANSI Fortran-77 standard [4]. In particular, absolutely no assembly code will be allowed within the program file, and no external programs may be referenced other than the standard Fortran functions. Fortran subprograms may be referenced only if the Fortran code for the subprograms is included in the program file and conforms to the other requirements mentioned in this paper. Finally, no modification to the algorithms in the code may change the number of floating-point operations performed.

The precision level of all floating-point data and operations in the program must be 64 bits, with at least 47 mantissa bits. As a test of the hardware precision, and to ensure that any modifications made to the program file have not fundamentally changed the calculations being performed, an accuracy check is included with each of the seven tests. These checks are performed by comparing a selected result from each of the programs with a reference value stored in the program code and then computing the fractional error. The total of the fractional errors from the seven programs must be less than 5×10^{-10} .

The NAS Kernel Benchmark Program automatically calculates performance statistics and outputs this report on Fortran unit 6. This report includes the results of the accuracy checks, the number of floating-point operations performed, the CPU run times, and the resulting MFLOPS rates. The total error, total floating-point operation count, total CPU time, and the overall MFLOPS rate are also included.

Normally only uniprocessor results are tabulated. If desired, multiprocessor performance may be estimated by simultaneously running the benchmark program on each of the individual processors. A multiprocessing performance figure may then be computed by averaging the timings from the runs on the individual processors. Although no explicit multiprocessing is performed in this manner, such an exercise measures the amount of interprocessor resource contention, which is a significant factor in multiprocessing. In this way the performance increase that can be expected from multiple processor computation can be estimated without making the laborious modifications that are usually required to invoke true multiprocessing.

REFERENCES

1. Hockney, R. W., and Jesshope, C. R., *Parallel Computers*, Adam Hilger, Bristol, England, 1981.
2. Brigham, E. Oran, *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, N.J., 1974.
3. Worlton, Jack, “Understanding Supercomputing Benchmarks”, Datamation, September 1, 1984, p. 121.
4. American National Standards Institute, *ANSI Fortran X3.9-1978*, ANSI, New York, 1978.

APPENDIX:

PROGRAM LISTING